

THE PROBLEM OF SELF-KNOWLEDGE

**Karsten R. Stueber
Department of Philosophy
College of the Holy Cross
Worcester, MA 01610
kstueber@holycross.edu**

Published in *Erkenntnis* 56 (2002), 269-296. Please quote according to the published version. If you need a hard copy of the article please feel free to contact me at kstueber@holycross.edu.

Abstract: This article develops a constitutive account of self-knowledge that is able to avoid certain shortcomings of the standard response to the perceived *prima facie* incompatibility between privileged self-knowledge and externalism. It argues that if one conceives of linguistic action as voluntary behavior in a minimal sense, one cannot conceive of belief content to be externalistically constituted without simultaneously assuming that the agent has knowledge of his beliefs. Accepting such a constitutive account of self-knowledge does not, however, preclude the conceptual possibility of being mistaken about one's mental states. Rather, self-knowledge has to be seen as only a general constraint or as the default assumption of interpreting somebody as a rational and intentional agent. This is compatible with the diagnosis of a localized lack of self-transparency.

Introduction

After the demise of the Cartesian picture of the mind and with it the predominance of the first person perspective and introspective psychology, few would define mental states as states to which we have indubitable, incorrigible and infallible access. Psychoanalysis and other psychological research have shown that we do not always have introspective knowledge about our own psychological states. One is even tempted to say that others may sometimes know my "mind" better than I know it myself.¹ Nevertheless, this does not imply that other people's knowledge of my mental state is usually on equal epistemic footing with my own knowledge. We still grant first person reports about one's own mental states privileged epistemological status. In contrast to such descriptions from the third person perspective, self ascriptions are supposedly based on "direct" and "immediate" access to one's own mental states, i.e. first person knowledge is non-inferential, not based on any evidence, and is not required to be justified in light of such evidence. A first person report is normally justified in being sincerely uttered.

Recently, philosophical controversy has arisen about whether first person authority can be satisfactorily explained in the context of our folk psychological discourse, which is constituted by seemingly inconsistent components. It has especially been felt that the central epistemological status of first person authority cannot be easily reconciled with an externalist account of meaning and

mental content that some philosophers have argued for on grounds of certain intuitions about attribution of content in the context of the well-known twin-earth thought experiments.

Such incompatibility between the assumption of self-knowledge and externalism could, of course, be easily resolved by giving up on either of them. One might withhold allegiance to externalism because it is "merely" a philosophical account of content and for that reason not as fundamental as our intuitions on self-knowledge and first person authority. Or one might argue against first person authority by stressing that one should rather dispense with such intuition since it is not supported by our best theories about meaning and content.²

In my opinion both of these "solutions" have serious drawbacks. In the end, they leave us with an unsatisfactory account of our folk-psychological notions of meaning and content, or so I would be prepared to argue. For the purposes of this paper it is more important, however, that neither solution is required. As I will show in the following pages, by focusing on knowledge of one's own beliefs, the assumption of self-knowledge and first person authority should be regarded as being constitutive for our conception of linguistic agency and our interpretive practices within the realm of folk-psychology, even if one regards meaning and content to be externalistically constituted. This constitutive account will also avoid certain shortcomings of the standard response to the perceived *prima facie* incompatibility between privileged self-knowledge and externalism. It is better able to meet the following constraints that I regard to be minimal for any theory accounting for the epistemological status of first person authority and self-knowledge within the folk-psychological context.

1.) **Asymmetry Constraint:** A theory of self-knowledge should be able to explain the fact that second-order beliefs about one's first order mental states count as real knowledge, although these second order states do not seem to require any further epistemic justification. They normally do not have to be based on any empirical evidence or inference like ascriptions from the third person perspective.³ To explicate self-knowledge requires taking into account this epistemic property, which I will refer to as its non-inferential and non-empirical character.⁴

2.) **Univocality Constraint:** Although first person and third person reports are based on different criteria, we still assume that the same mental concepts are used. Otherwise we could not use self-ascriptions of others in trying to understand them as we normally do. A theory of self-knowledge has to explain how this assumption can be justified.

3.) **Constraint of Semantic Externalism:** As Putnam, Burge, Davidson and others have made plausible, we base our attribution of intentional states on the external causal relations between a person and his or her environment.⁵ Any account of self-knowledge therefore has to be compatible with this philosophical analysis of content. Specific attention has to be paid to the prima facie incompatibility of the asymmetry and externalism constraint.

4.) **De dicto Constraint:** Within the folk psychological context, the attribution of de dicto beliefs tend to play a central role in our attempt to rationalize the behavior of an agent, because they are thought of as expressing how agents conceive of their world. These de dicto beliefs are also expected to play a role in the rational deliberation and decision process of an agent based on her knowledge of the world and her mind. Without the assumption of self-knowledge it would, for example, be inexplicable why an agent should rationally decide to go to the kitchen and drink a glass of water if she does not know that she believes that there is a glass of water in the kitchen. De re attributions of the form "she believes of water that it is in the kitchen" however do not play this role in the explanation of behavior and neither are they expected to play any role in the deliberation procedure of the agent. A theory of self-knowledge which cannot account for the de dicto/de re distinction in this respect should be regarded as insufficient.⁶ Note, however, that in formulating the *de dicto* constraint in this manner, I am not denying that thoughts containing indexicals, especially first person indexicals, are essential for the explanation of certain actions.⁷ My remarks are intended to be limited to the attribution of fully conceptualized thoughts, which do not contain indexicals.

5.) **Fallibility Constraint:** A theory of self-knowledge has to be compatible with the logical possibility of error and self-deception.

Although I do not have the space here to argue for these assumptions extensively, I take each of them to be sufficiently supported by philosophical arguments and the explanatory and interpretive practices of folk-psychology. For that reason these constraints will constitute the argumentative background assumptions for my constitutive account of self-knowledge.

My argument will proceed as follows. In the first section I will mainly explicate the structure of the skeptical argument that leads one to conclude that externalism and self-knowledge are incompatible. In the next section, I will outline the standard response to the skeptical challenge and will explain why I regard it to fall short of fully accounting for self-knowledge and for the central status of first person authority in the folk psychological context. The standard strategy is unsatisfactory, because it deals with the question of the compatibility of self-knowledge and externalism without investigating how external factors enter into the determination of belief-content in the interpretation of another speaker. Yet if one pays attention to these facts and if one conceives of linguistic behavior as voluntary action in a minimal sense, then one cannot -as will be explained in the third section- conceive of belief content to be externally constituted without self-knowledge.

Accepting such a constitutive account of self-knowledge does not, however, preclude the conceptual possibility of being mistaken about one's mental states. Rather, as will be indicated in the last section, self-knowledge has to be seen as only a general constraint or as the default assumption of interpreting somebody as a rational and intentional agent. This is compatible with a diagnosis of a localized lack of self-transparency in situations in which we otherwise cannot understand the agent's behavior as a rational action.

Self-Knowledge and The Externalist Challenge

An intuitively appealing way to explain the possibility of self-knowledge is to postulate some inner mechanism of perception, which provides information about our own mental states. Just as we gain knowledge about objects of the external world through certain sensory mechanisms without relying on further inferences, we gain knowledge about our own mental states by getting some information about certain inner and inherently private objects.

This object-perception model of self-knowledge, however, has to be rejected, if one accepts the above constraints on a theory of self-knowledge. In light of Wittgenstein's remarks against the Cartesian model of the mind, one could argue that the object-perception model is unable to meet the univocality constraint. Such a conception of inner perception can also not account for the non-inferential nature of self-knowledge, given externalism. If the identity of my "water" thought depends on my being in an H₂O environment and not in an XYZ context, then the identification of an inner object as a water or twater thought would require knowledge about the relational feature of that particular object. Knowledge of the content of one's own thought would seem to be inferred from knowledge of the external world.⁸

As is well known, Wittgenstein abandoned any attempt to account for self-knowledge on the basis of a particular mechanism of inner perception and opted for a version of a constitutive account of self-knowledge. According to his rule-following considerations, no inner state can account for its representational property because the representational properties of language and thought cannot be understood as being logically independent of a specific social or individual practice. For Wittgenstein, the fact that self-ascriptions of mental states have a special status within our normal linguistic practice cannot be explained through a specific cognitive mechanism. It is merely a grammatical fact of our mental discourse.

While I am certainly sympathetic to such a Wittgensteinian move, Wittgenstein's rather dogmatic declaration of self-knowledge as a grammatical fact of our language game is

philosophically unsatisfactory. It is especially not clear how the declaration of the constitutivity of self-knowledge for the intentional discourse can philosophically answer certain skeptical doubts about our ability to know non-empirically and non-inferentially the content of our own beliefs that seem to arise if one accepts externalism.⁹

The skeptic claims that if we accept semantic externalism and Putnam's twin earth intuitions then one has to explain why my second order belief that I am thinking that water is tasteless (i.e. my second order belief that p) can count as knowing that I am thinking that water is tasteless, since I do not seem to be able to exclude non-inferentially the skeptical hypothesis that I am in fact thinking that twater is tasteless (i.e. q). According to the intuitions regarding travel between earth and twin-earth I could be switched to twin earth without even recognizing introspectively and non-empirically that a switch has taken place. Without me noticing, my daily water thoughts and conversations about water would slowly change into twater thoughts and conversations about twater. If these intuitions are plausible, then q is in this context a relevant alternative whose truth is incompatible with p. In order to be credited knowledge about p one has to know that q is false. Therefore, to know that I am thinking that water is tasteless I have to know that I am not thinking that twater is tasteless. But I cannot do that without having further empirical knowledge about the world I am in. The assumption of non-inferential self-knowledge, thus, seems to be without any justification in the context of semantic externalism. According to Brueckner,¹⁰ one can represent the skeptical challenge in the following way. I will refer to it as the skeptical argument (SA):

- i.) If I know that I am thinking that water is tasteless then I know that I am not thinking that twater is tasteless.
- ii.) I do not know that I am not thinking that twater is tasteless.
-
- iii.) I do not know that I am thinking that water is tasteless.

Premise (i.) follows from the principle that knowledge is closed under known entailment and premise (ii.) has to be accepted because of the above considerations about twin earth travel. In this context, Wittgenstein's grammatical account of self-knowledge and his claim that philosophical

problems arise only because of the misuse of ordinary language cut no ice. Even within our ordinary linguistic practice, we are faced with a dilemma of how to reconcile semantic externalism with central aspects of our concept of knowledge.

II

The Standard Strategy of Reconciliation and its Problems

The standard strategy¹¹ to meet the above challenge is to argue that the reasoning for the incompatibility of self-knowledge and externalism is fundamentally flawed and rests on a logical confusion. For the externalist, it is constitutive that somebody stands in a certain causal relation to his or her environment in order for that person to possess a mental state with a specific content. But externalism is not committed to claim that one has to know that these conditions obtain in order to be in a particular mental state or to be introspectively aware that one is in this particular state.

The paradox of externalism then arises, or so it is claimed, because of the acceptance of the invalid principle that in order to be aware that one thinks that *p*, one has to know the constitutive facts that individuate the content of that thought. All that externalism requires for having a particular thought is that the external factors and conditions that are constitutive for the individuation of a specific thought content do obtain. By the same externalist reasoning all that is necessary for having a second order thought is that its individuating conditions obtain.

This however can only be regarded as a first step in answering SA since it does not directly address the question how externalism is compatible with my knowing that I have certain first order thoughts with a certain content. It only shows that externalism can be reconciled with my being able to think or to believe that I am thinking that water is tasteless, but it does not directly show that my second order beliefs constitute knowledge of my first order thoughts.¹²

In order to answer this worry, Burge focuses originally on the "self referential and self-verifying" character of so called Cartesian thoughts, i.e. thoughts or judgment which concern one's own occurrent first order thoughts in which I entertain a certain proposition.¹³ If you think that I am

thinking that it is a nice day outside you do not have any epistemic guarantee that I have such a first order thought. You would need further evidence in order for your belief about my mental states to count as knowledge. I, on the other hand, in having such a second order thought do at the same time entertain the thought that it is a nice day outside. The content of my first order thought is thus automatically part of my second order thought. Regarding Cartesian thoughts the question of making a mistake cannot really arise because all introspective second order thoughts have to be true in order for us to have them.

For Burge, the recognition of the self-verifying character of Cartesian thoughts implies that one has to count one's own second order thoughts as self-knowledge in the strictest sense. Regardless of whether I am on Earth or Twin-Earth I will have true second order thoughts about thinking that water or twater is tasteless without being able to decide non-inferentially whether a switch has occurred.

Although I will argue that one has to agree with this judgment in the end, the argument itself does not directly answer the prima facie plausibility of SA. One might therefore claim that the above argument has only shown "that the act of thinking a thought makes the thought true" but that this does not entail "that what is thought is known to be true."¹⁴ In order to quell these misgivings one has to focus in more detail on the non-inferential, non-discursive and non-empirical character of self-knowledge.

The skeptical argument apparently works because the first premise seems to be derived from the closure principle. The skeptic argues for the second premise (i.e. I do not know that I am not thinking that twater is tasteless.) by pointing out that I do not have any introspective and non-empirical knowledge about whether or not I live on Twin-earth. But this requires that the type of knowledge referred to in the second premise and in the first premise has to be non-empirical knowledge because I do know or at least could know empirically that I am living on earth. I only have to have evidence for the claim that water is H₂O. SA* (the explicit version of SA) has thus the following form:

- i.) If I know non-empirically that I am thinking that water is tasteless then I know non-empirically that I am not thinking that twater is tasteless.
- ii.) I do not know non-empirically that I am not thinking that twater is tasteless.

- iii.) I do not know non-empirically that I am thinking that water is tasteless.

Premise (i) could be derived from the following closure principle: If S knows non-empirically that p and S knows non-empirically that (p > not q) then S knows non-empirically that not q. However, in the context of the twin-earth example in order to know that p (I am thinking that water is tasteless) entails not q (I am not thinking that twater is tasteless), requires that I have to have both water and twater concepts and that I know that they are not identical. But that is possible only if I know that water and twater have different chemical structures. It is something I cannot know non-empirically. The first premise therefore does not have to be accepted because the special circumstances of the twin earth thought experiment do not allow us to apply the modified closure principle in this case. My not knowing non-empirically that I am not thinking that twater is tasteless is, thus, not logically incompatible with my knowing non-empirically and non-inferentially that I am thinking that water is tasteless.¹⁵

However, insofar as the standard strategy emphasizes Cartesian thoughts it is rather limited as a general account of the presumption of self-knowledge. Not only are we supposed to know our occurrent first order thoughts but also those thoughts which have occurred in the immediate past, standing beliefs, desires, fears etc. But believing that we fear snakes or believing that we believe that Santa Claus exists does not logically imply that we actually do have these first order fears and beliefs. The first step of the standard argument only shows that our prima facie misgivings concerning the incompatibility of externalism and self-knowledge about the content of our mental states might be misconceived but it does not explain why the assumption of self-knowledge is central within the folk psychological context. What we still need is a justification for the fact why our second order beliefs should in general count as knowledge. Correlated with this general problem, there are two more specific arguments that further confine the effectiveness of the standard response

to externalist worries:

1.) Boghossian claims that even if one accepts that externalism allows for knowledge of one's occurrent thoughts it cannot explain the assumption of epistemic transparency of de dicto thoughts. It cannot explain how we assume that we know **non-empirically** whether two thoughts have the same or different content. This assumption plays a central role in the evaluation of the rationality of an agent. Because only under the assumption that the person knows that the sentence "water is not tasteless" contradicts the sentence "water is tasteless" can we justifiably judge her to be irrational if she expresses both beliefs. Otherwise her "irrationality" is excused in the same way we excuse somebody who utters the English sentence "water is tasteless" and the German sentence "Wasser ist geschmacklos" without knowing the meaning of the German sentence. According to Boghossian, thought experiments about twin earth travel challenge the transparency assumption in the following manner: Imagine that somebody is switched to twin earth and continues living there. Then under the assumption that "beliefs about the past... will retain their earthly interpretation," she will not recognize that her past water beliefs and her recent twater thoughts are different.¹⁶ Therefore, she might not recognize that the belief that water is tasty which she had as a child does not contradict her adult and more mature belief that twater is tasteless.

2.)The standard strategy only addresses the question of how externalism can be reconciled with our knowledge of the content of a propositional attitude, since the skeptical argument emphasizes this problem. But in our ordinary conception of self-knowledge we assume that we know both the content and the attitudinal component of our mental states. I do not only know that my thoughts concern the proposition that snakes are dangerous but whether or not I believe it, fear it or get excited about it. The standard strategy makes clear that externalism does not necessarily create a specific problem for the knowledge of the content of my thoughts but seems to allow for the conceptual possibility that I am mistaken about the attitudinal component.¹⁷

As I will show in the next section, if one pays close attention to our practice of attributing intentional states and our conception of linguistic agency, it will become clear that we are essentially committed both to view the *de dicto* content of our mental states - especially beliefs - to be externalistically constituted and to assume that linguistic agents have access to their mental states. This will provide the needed justification for why our second order beliefs should count as knowledge, i.e. it will supply us with the required second step of answering the skeptic. Within this context it is also possible to answer the questions regarding the transparency of mental content and the attitudinal component of self-knowledge.

III

A Constitutive Account of Self-Knowledge

In the following I will show that self-knowledge should be viewed as an implicit and essential assumption of the attribution of intentional states to a linguistic agent, at least if we conceive of linguistic action as a voluntary activity and linguistic communication as a cooperative enterprise. In the context of such a conception, the assumption of self-knowledge is justified as it provides the best explication for the possibility of the interpretative process, which is constituted by certain interpretative principles. In this sense, I conceive of my explication of self-knowledge as a constitutive account.

My account of self-knowledge is sympathetic to Davidson's remark that without the presumption of self-knowledge "there would be nothing for an interpreter to interpret" and his claim that "the agent herself ... is not in a position to wonder whether she is generally using her own words to apply to the right objects and events."¹⁸ However, it is difficult to find in Davidson any explicit argument for the claim that one needs to attribute true second order beliefs to speakers in order to conceive of them as interpretable. Even if one agrees with Davidson's analysis of radical interpretation and his insistence that in interpreting somebody we have normally to assume that the

speaker expresses mostly true beliefs, one might wonder why it is also necessary to assume that she has true second order beliefs and why it is required to regard such second order beliefs as constituting knowledge. Yet such a demand for explication should not be understood to depend on accepting the metaphor of "objects before the mind," which Davidson correctly rejects. It is better understood as a question of conceptual clarification depending on our ordinary understanding of the concept of knowledge, according to which it is minimally required for somebody to know that p that she has to believe that p. To say it more succinctly, what is missing in Davidson is an argument for the claim that our practical knowledge of meaning -i.e. know how- manifested in our ability to use language correctly implies *knowledge that* we have certain beliefs.¹⁹

Prima facie Shoemaker's attempt to show that it is of the "essence of the mind...that each person has a special access to his own mental state"²⁰ seems to be of some help in this regard, especially if one keeps in mind that, for Davidson, the interpretation of a speaker can only proceed under the assumption that she is a rational person. Shoemaker argues that self-knowledge and second order beliefs are supervenient on first order beliefs and desires "plus a certain degree of rationality, intelligence and conceptual capacity,"²¹ since it is impossible to coherently conceive of a rational but self-blind person. A self-blind person is somebody, who would have our intelligence and conceptual capacity but who lacks a special first person access to her own mental states. She would be able to attribute mental states to herself only in a third person manner, based on her knowledge of her own behavior. For Shoemaker, however, such self-blindness could not be attributed to another person because that person behaves "in ways that provide the best possible evidence that she is aware of her own beliefs and desires to the same extent that a normal person would be, and so is not self-blind."²² Shoemaker emphasizes, for example, that such a person would recognize the logical impropriety of asserting "I do not believe that p" when she asserts that p, assuming that she recognizes that an assertion is the expression of a belief. She thus would always assert that "I believe that p" whenever she asserts that p. Such an assertion is the expression of a second order belief.

Nevertheless, one has to be careful. All that Shoemaker proves is that the rational person behaves as if she has self-knowledge in avoiding Moore-paradoxical utterances. But this does not prove that she behaves in that manner because she has self-knowledge. Only the second claim would really establish that immediate first person access is essential for having a mind. So far it has not been established that self-knowledge is required for the making of an assertion. It is equally plausible to maintain that a self-blind person infers her belief that *p* by asserting *p* in the same manner as the hearer of the utterance infers such belief. As Gareth Evans suggests,²³ we find out about whether we believe that *p* not by looking inward but by directing our gaze at the world and by deciding whether we would assert that *p*. In order to show that self-knowledge is constitutive for having a mind, one would have to argue that having a mind cannot be explicated without direct first person knowledge. It is not sufficient to show that having certain mental capacities implies that one acts as if one has self-knowledge.

It is therefore necessary to pursue another line of argument. As a first step, it is helpful to recognize the central role of speakers' intentions in explicating our concept of a successful communication. Using language is not a form of instinctive behavior but it is an action under the control of a specific agent. As is the case in regard to non-linguistic actions, we understand a linguistic action only by grasping the intentions that caused the agent to behave in a certain way. Only in regard to such intentions can we also distinguish between different actions that are implemented by the same physical behavior. One's raising of a hand can, for example, either be the indication that one wants say something or the attempt to point to something in the ceiling, depending on what one intended to accomplish with one's behavior. Similarly, in order to fully understand the utterances of a speaker we not only have to recognize their content we also have to know for what illocutionary purposes she uses the uttered sentences.

Speech acts can be distinguished by very special kinds of intentions,²⁴ namely intentions to express or reveal certain mental attitudes or mental states, *whereby an intention to express an attitude or mental state that p is understood as the intention to make one's audience recognize that*

one has the attitude that p. The utterance "Is the door shut?" can for example be understood as a question or as a request, depending on whether the speaker intends that utterance as the expression of her desire for knowledge or of her desire for a certain action on the part of the hearer. Similarly, the utterance of the sentence "The snow is white" could be a mere attempt to learn the correct pronunciation of the English sentence or the recital of the first line of a Christmas poem. It is an assertion that the snow is white only if the utterance is intended as the expression of the corresponding belief. According to this account, the speech act of an assertion is the intentional verbal expression of a belief because the utterance is caused by an intention to express the belief. For that very reason, a speaker would find it puzzling and would take it as a sign of miscommunication, if after her assertion of that *p* she would be asked whether she believes or thinks that *p*.

An assertion is thus not merely the expression of a belief in the sense of an utterance being caused directly by the belief itself.²⁵ Such an explication of an assertion would have the counterintuitive consequence that only a sincere assertion can count as an assertion, since only in this case could a belief actually cause an assertion. It would mean that a liar does not really assert anything, even though the success of her lie depends on the fact that her utterances are mistaken as genuine assertions. But to be able to regard both an insincere and a sincere assertion as the same type of linguistic act would require that one understands them both and in the same manner as expressions of beliefs. To allow for the possibility of insincerity in the performance of a genuine speech act implies that one has to understand the speech act not as being caused by the mental state it is expressing. One has to conceive of it as being caused by the intention to express a certain mental state, since one can form such an intention, even if one does not actually possess the mental state one intends to express.

If the above conception of language use is plausible, then we interpret a speaker correctly only if we grasp the mental attitude the speaker intended to express.²⁶ The challenge for my constitutive account of self-knowledge is to show how such a conception of communication entails

that the speaker has to have true second order beliefs that should be counted as knowledge of her first order beliefs.²⁷ If an assertion requires only an intention to express a certain belief why should we furthermore assume that an assertion also requires true second order beliefs or at least the capacity to form true second order beliefs?

In order to answer this question, and as a second step in my argumentative strategy, it is necessary to have a closer look at the constitutive principles and assumptions of our interpretative practices, as they are revealed by an analysis of radical interpretation. At the start of radical interpretation, the radical interpreter has to assume that the speaker expresses sentences, which she holds to be true and that are true. Otherwise one would not be justified in taking the environment, in which "observation-sentences" such as "Da ist ein Baum" are uttered, as providing clues for developing one's interpretative hypotheses. One thus has to assume that the speaker makes assertive utterances and that she intends to express certain of her beliefs. To be more precise, it is a constitutive assumption of radical interpretation that the speaker utters a sentence *p* in a context in which *p* is true, with the intention to express her belief that *p*, and at a time when she actually believes that *p*, i.e. her assertions are sincere.

The above assumption about the sincerity of the speaker does not automatically imply that we also have to presume that the speaker has true second order beliefs about the beliefs she intends to express. It is only required that she, as a matter of fact, asserts sentences that correspond to her beliefs. I would however argue that the assumption of true second order beliefs is implied by the above constitutive interpretive assumptions, if we also conceive of a linguistic utterance as a voluntary action in a minimal sense. In conceiving of linguistic acts as voluntary acts in a minimal sense we assume that they are under the control of the agent, i.e. *she has a choice between asserting something sincerely or insincerely*.²⁸ One therefore has to ask what reason do we have for assuming that the speaker is sincere and only utters sentences that are true and that she believes to be true. Since linguistic action is assumed to be voluntary in the above sense, we know that it is not the case

that such a relation between speaker and environment holds because of a strict natural law or because the utterances of a human speaker are automatically triggered by certain environmental cues. If the speaker so chooses, she could, for instance, refuse to cooperate and try to mislead the interpreter about her beliefs in order to laugh at the interpreter's stupidity.

In the context of such a conception of language use, an interpreter - at least implicitly - has to assume that the speaker cooperates in the communicative enterprise because she intends not to mislead the interpreter about the belief she is expressing. One has also to presuppose that it is not pure chance that she is successful in implementing that intention. But in order to make sense of these assumptions, one has at least to assume that the speaker has true second order beliefs about what she believes, because otherwise we would have no reason to think that she is able to successfully act according to her intention not to mislead the interpreter and to express the belief that *p* when she actually believes that *p* is true. If one does not have a true second order belief about one's first order beliefs one could also not have a belief about whether or not a specific utterance is misleading in a certain situation. In order to know that, one has to be able to distinguish the situations in which one intentionally expresses a belief that one has from those in which this is not the case. Only because I know or because I have true second order beliefs about what I believe in a certain situation, can one reasonably expect that I intend to express a certain belief that does not mislead the interpreter. To assume the existence of true second order beliefs is therefore implicit in the constitutive assumptions which guide our interpretative enterprise, at least if one assumes that the uttering of true sentences is not an automatic process determined by laws of nature. It allows us to provide a plausible account of the fact that the assumption of sincerity is true of a particular speaker, given that linguistic action has to be regarded as a voluntary action.

But is this line of thought simply begging the question against the skeptic? The skeptic might question whether the appeal to true second order beliefs is the only plausible explication of the speaker's behavior in the situation of radical interpretation. Could it not be the case, one might argue, that the speaker intends not to mislead the interpreter, but that she has only false second order

beliefs. She forms thus the intention to express the belief that p and asserts that p, even though she does not believe that p. This would be a case of unintentional global insincerity. Or to take a second scenario, without her knowing and while having such false second order beliefs, she expresses her actual belief that q by uttering q. The skeptic would therefore request an assurance that the above divergence of first order and second order belief does not exist.

In response to such concerns it has to be stressed that the assumption of true second order beliefs is necessary only under the assumption that speech behavior is voluntary. It is logically possible that somebody behaves sincerely and in a manner that is interpretable from the perspective of the radical interpreter without having any second order beliefs. Indeed such a person could furthermore be characterized as having no intention to mislead the interpreter (even though he cannot have an intention not to mislead.) However, such behavior could hardly be regarded as voluntary activity, since for such a person a lie is not even a conceptual possibility. If one makes this commonsensical assumption about the nature of speech behavior then the above worries are either empty misgivings, since they can never be substantiated within our interpretative practices, or they are incompatible with the above conception of speech behavior. Take the first case. While such unknowing insincerity is possible in a limited number of cases, to assume it on a global level is more than implausible. In order for the skeptic's worries to have any force, the skeptic would have to maintain that such linguistic behavior would be indistinguishable from the behavior of the self-knowing speaker who intentionally expresses the beliefs she holds to be true. This thesis is equivalent to the claim that such a self-deceived speaker is able to provide consistent clues for the interpretation of her speech without having any true first order beliefs about the world. But such an assumption would make it difficult to explain how the speaker in question can behave in a manner vis a vis the environment that is indistinguishable from the speaker who has true beliefs about it. Whatever first order belief such a person might have, they do not seem to fulfill any causal role we associate with the notion of a belief. The second scenario is, on the other hand, equivalent to the assumption that the speaker utters a sentence that she holds to be true if and only if it is true because

of a law of nature. The assumption that one has only false second order beliefs but that it does not change one's linguistic behavior in the situation of radical interpretation is equivalent to the assumption that there are no second order beliefs. Whereas I admit that this is a logical possibility, it does not square with the conception of linguistic action as voluntary behavior.

One should notice that the above explication of the central status of self-knowledge within the folk-psychological context applies both to knowledge of content and to the attitudinal component of one's belief state. In attributing to somebody the intention to express her belief in a non-misleading fashion, we cannot assume that the person knows only the content of her belief but might be mistaken that it is a belief. One can only intend to assert *p* sincerely insofar as one intends one's utterance to be an expression of one's belief that *p*. Knowledge of the attitudinal component is thus as much a constitutive principle guiding the interpretation of a specific person as is knowledge of content of one's own mental state. Similar remarks apply to Boghossian's worries that externalism is in general not compatible with the assumption of the semantic transparency of our own thoughts. He might argue that the argument so far only shows that we have knowledge of a specific thought but it does not illustrate that we also have knowledge of whether two thoughts have the same or different content. However, a short consideration of the holistic constitution of *de dicto* content should reveal that the assumption of linguistic agency requires that we also assume that our thoughts are to some extent transparent to us. As I will argue in the next section, Boghossian's thought experiments show only that we lack transparency some times. They do not prove that we do not know for any two thoughts whether or not they have the same content. Indeed, the thesis of meaning holism implies that one can only justify the interpretation of one sentence in the context of interpreting a number of other sentences. This requires that we assume that the agent on the whole subscribes to the same normative rules of rationality and consistency as we do.²⁹ To be able to assume that the speaker intends her linguistic intention to be recognized then also compels one to presume that the speaker will intend to use her linguistic expression in a way that is consistent with her earlier and other speech acts. She should be generally aware of any inconsistencies in her belief system. Such

awareness, however, presupposes the ability to compare the content of different thoughts and beliefs. We therefore cannot assume non-empirical knowledge of content of a belief without a certain capacity to compare it to the content of other beliefs.

Even if we agree with the argument so far, the above considerations seem to raise two important objections. First, how does this account, which appeals to intentions as the criterion for the correctness of an interpretation square with an externalist account of content? And secondly, how does such an account answer skeptical doubts regarding self-knowledge, especially the question of what allows us to regard true second order beliefs as knowledge?

Regarding the first issue, one might be concerned with the fact that the above account implicitly relies on an internalist notion of content, since the speaker's intention and not facts which are necessarily accessible to a third person perspective seem to constitute the criterion for the correctness of an interpretation. To answer these concerns one should point out that nothing so far assumes that intentional content is internalistically constituted, since as the first step of the standard argument points out, externalism is not incompatible with self-knowledge. Moreover, if one regards - as I do - the analysis³⁰ of radical interpretation as an analysis of the factors that are constitutive for somebody having contentful mental states, then radical interpretation has to proceed under the assumption of self-knowledge and an externalistic notion of content. From the point of radical interpretation, in order to conceive of the speaker as having a mind (and certain intentions) we have to assume that the speaker stands in a systematic relationship to her environment. The radical interpreter proceeds only under the assumption that the speaker has a mind and that she is a part of these constitutive systematic relations to the world. Only then is the interpreter justified to take the observed interaction between speaker and the world as evidence for her interpretation.³¹

Equally important are Brueckner's worries whether true second order beliefs really amount to instances of non-empirical knowledge. In response, one should remind the skeptic of the result of the second section of this article. The skeptical argument (SA*) does not force us to deny the possibility of non-empirical self-knowledge, even though I do not know non-empirically that I am

not thinking that twater is tasteless. Insofar as content is externally constituted one has to interpret one's twin as talking about twater and as having a true second order belief that she is talking about twater without requiring that she can distinguish non-empirically between water and twater thoughts. The skeptical argument thus fails to show positively that we lack self-knowledge and that we cannot regard true second order beliefs as self-knowledge. And, even though true beliefs without epistemic justification do not normally count as knowledge, the above considerations should be seen as providing the necessary epistemic justification for conceiving of second order beliefs as being in general truth-conducive. They thus provide an argument for the claim that such second order beliefs constitute self-knowledge. It answers the skeptical challenge to the possibility of self-knowledge by pointing out that insofar as the skeptic thinks of herself and others as interpretable linguistic agents she already is forced to grant what she is challenging. But this is an assumption that the skeptic commits herself to by posing any skeptical challenge, since she otherwise seems to deny the intelligibility of the skeptical challenge itself. The above argument thereby rejects any epistemic reason to globally doubt the truth-conduciveness of second order beliefs as being intelligible. For that very reason, self-reports have normally to be regarded as being epistemically justified if the speaker sincerely makes them, because this is implied in our practice of treating each other as linguistic agents. Our epistemic right to self-knowledge is hence grounded in our linguistic agency.

Notice, however, that the above considerations provide an argument only against the global challenge to the possibility of self-knowledge. It should be seen as an argument for the default assumption of self-knowledge insofar as linguistic agents are concerned. Nothing that I have said, as will become clear in the last section of this article, excludes the possibility that a speaker errs about some of her first order mental states or beliefs. Yet such challenges to particular self-reports have to be argued for in light of specific evidence, for example, that what the speaker says now contradicts what she said five minutes ago.

Before I address this issue in more detail, a few remarks about the difference between my account and Burge's recent account of the epistemic status of the assumption of self-knowledge are

in order.³² I share with Burge the general strategy of deriving our justification to self-knowledge from a fact about ourselves that seems to be undeniable even from the skeptic's perspective. I also agree with Burge that not every person who has self-knowledge would be able to articulate such justification. The above argument constitutes an epistemic justification in the sense of Burge's notion of "entitlement," i.e. as "epistemic rights or warrants that need not be understood by or even accessible to the subject" of this epistemic warrant.³³ I am however not so sure that Burge locates the source of this entitlement in a condition that cannot be denied by the skeptic. He derives our "entitlement to self-knowledge" from our self-conception as rational and critical reasoners. Our capacity to critically reflect on our thought processes and ability to evaluate whether our thoughts conform to rational standards would be impossible to conceive unless we presuppose direct self-knowledge about our own mental states. Without such an assumption one could not fathom how we could have such reflective capacity and one could not explicate how such reflective capacity could provide rational grounds for the revision of our first order thoughts if they are found rationally to be inadequate. As Shoemaker, in another argument for the assumption of self-knowledge suggests, second order beliefs and desires rationalize the revision within the system of first order beliefs in light of new experience.³⁴ While I do not deny that our conception as reflective reasoners implies the assumption of self-knowledge, it is in my opinion not clear that it constitutes the source of our entitlement to self-knowledge, especially in light of Boghossian's challenge which I referred to in the last section. For Boghossian, semantic externalism was incompatible with the assumption of the transparency of de dicto thought and hence incompatible with conception of us as critical reasoners. To insist that we conceive of ourselves as critical reasoners and therefore as persons with self-knowledge would not solve this problem. It would rather lead to a paradox. On the one hand we have to conceive of ourselves as having self-knowledge, because we are critical reasoners. On the other hand our assurance that we have self-knowledge and are therefore critical reasoners can be independently challenged.³⁵

My explication of the epistemic status of self-knowledge avoids these difficulties. By

emphasizing the concept of linguistic agency in the context of radical interpretation, I derive rationality and self-knowledge simultaneously as constitutive assumptions of the process of interpretation. The source for our entitlement to self-knowledge and rationality lies in the fact that we cannot claim that we or another person have any contentful mental states unless we assume that we have self-knowledge and that we are rational.

IV

The Fallibility of Self-Knowledge

My constitutive account of self-knowledge should not be misunderstood. I am not rejecting Boghossian's intuitions that I am not non-empirically aware of the difference in content between my present twater and my former water beliefs. Indeed in order to be aware of such a difference, one could argue that one would need empirical knowledge regarding the chemical structure of these substances. I am just denying that one can derive a general conclusion about the incompatibility of the transparency of content and externalism from these considerations. Boghossian's argument is only successful against an account of self-knowledge in the context of an externalist theory of content, which does not allow for any error in one's self-conceptions.

Correctly understood, Boghossian's thought experiments reveal only that one should not think of our capacity to remember the past as providing us always with the capacity to re-experience the past in terms of the exact conceptual repertoire that we possessed at that time. But this is certainly not a assumption that we make in our ordinary folk-psychological practices. It is an even more problematic assumption in the more disciplined interpretive practices of the human sciences.³⁶ Here we ordinarily allow for the possibility that the autobiographical memories of persons are biased and filtered according to certain personal preferences, even without attributing any explicit intention to deceive. Especially after a long period of time encompassing a change of attitude and theory, we have to reckon with the likelihood - as any interpreter of autobiographical writing does - that the

agent interprets her past actions in light of her new theories, without necessarily recognizing the change of opinion she has undergone. For that reason, I would argue that one could accept Boghossian's twin earth intuitions without accepting his general conclusion. His considerations do not at all challenge our ordinary conception of self-knowledge, which allows for a certain amount of fallibility. I would rather suggest that we understand the scenario of twin earth travel in analogy to the limited authority we have in interpreting and remembering our past. After a certain period of time, probably the time it takes to change our water concepts into twater concepts, a misinterpretation of one's thoughts does not seem to be less plausible than a "misreading" by any person of her earlier life in light of a changed theoretical framework. Although I regard self-knowledge as a constraint that is essential to our conception of intentionality and meaning, my position does not imply that we can never assume a lack of self-knowledge in particular, localized circumstances. In this respect, the assumption of self-knowledge is similar to the presupposition of truth as a constitutive principle for any empirical interpretation. We have to attribute true beliefs to a speaker in order to understand her at all, yet this does not require us to attribute to her only true beliefs. In fact, we are sometimes forced to attribute false beliefs in order to explain differences between us and the observed behavior of the speaker and to understand the latter as rational behavior. Analogously, the attribution of a lack of self-knowledge is another way to fit an agent's "unexpected" and obviously less than optimal rational behavior into the framework of intentional explanation. In this manner we can "explain" prima facie inconsistent behavior in light of the agent's incorrect view about her own mental states. These explanations "work" because of our conception of what is minimally required for one's deliberations -either about one's future actions or how to change one's beliefs in view of recalcitrant evidence-to be rational in an optimal sense. For such deliberations to be optimally rational, it at least seems to be required that one has a correct view of the facts and a knowledge of one's desires and beliefs. If self-knowledge is required for optimally rational behavior, some instances of "irrational" behavior can be accounted for by attributing a lack of self-knowledge.

Take the example of wishful thinking. A mother might hold onto the belief that her son is a law-abiding citizen even though she repeatedly finds drugs in his bedroom. Her wish to save the harmony of family life and the wish to hold on to her idealized picture of the son is obviously the cause for her holding onto a false belief despite evidence to the contrary. This irrational behavior becomes "intelligible" if we consider the case from her perspective, assuming that her mind is not fully transparent to herself. She, for example, acknowledges the evidence but she does not regard it as overwhelming. She also tries to find reasons to disregard the evidence like his otherwise meticulous behavior etc.³⁷ Obviously, she is aware of her wish to hold on to a positive impression of her son but she is not aware that this wish actually causes her to disregard objective and overwhelming evidence. Therefore, even though it is objectively a clear case of non-optimal belief formation, the attribution of a certain lack of self-knowledge lets us explicate the mother's conclusions. Accounting for cases of less than optimal rational behavior in this manner, however, in no way diminishes our above conclusion regarding the constitutivity of self-knowledge in the folk-psychological context. Rather these explanations are possible because lack of self-knowledge has to be the exception, since otherwise we could not attribute any intentional states to a particular person.

Conclusion

As I have argued, the skeptical challenge to self-knowledge in the context of semantic externalism cannot be sustained. Firstly, the skeptic does not recognize that the non-empirical character of the knowledge of one's own thoughts does not require that one is able to exclude the respective twin earth thought. But, more importantly, the skeptic's worries are unfounded because the attribution of intentional states to a linguistic agent can proceed only under the assumption that the mental states are in general known to that person, even if psychological content is externalistically constituted. As a closer look at our interpretative practices revealed, causal relations

between the speaker and the environment can be taken as evidence for the individuation of content only under the assumption that the speaker has direct and immediate access to the content of her mind.

My proposal for a constitutive account of self-knowledge did not argue the claim that this is the only way of accounting for self-knowledge and first person authority and I do not want to unnecessarily burden my account with such a claim. Indeed, my proposal is in principle compatible with an account of self-knowledge such as Armstrong's, according to which our capacity to know our own thoughts is underwritten also by an empirical mechanism of the brain that allows us to "scan" our first order mental states. Armstrong conceives of scanning as a process of getting information about what particular mental state one is in by virtue of a reliable empirical mechanism.³⁸ According to this model, we know that we think that the sun is shining because that thought causes us normally to think that we think that the sun is shining.

However, I am skeptical that such a "naturalistic" account of self-knowledge could succeed since it would presuppose a reductive account of semantic *de dicto* content in terms of non-semantic categories. Only if a naturalistic account of content is possible, can we conceive of self-knowledge as being based on an empirically reliable scanning mechanism of the brain. But if Davidson, Putnam and others are correct, our normal attribution of *de dicto* content contains an irreducible, pragmatic element, insofar as belief attribution does not mirror exactly functional roles or causal relations to the environment.³⁹ Functional roles and causal external relations are taken into account in the interpretation of another speaker but the final criteria for a successful interpretation is the best fit between the belief system of the speaker and that of the interpreter. Such a fit however cannot be characterized functionally.

In view of the implausibility of a reductive account of folk psychological content, my account of self-knowledge is the more plausible alternative of accounting for self-knowledge. The assumption of self-knowledge is best viewed as an integral and constitutive part of the intentional framework, which alone allows us to fully comprehend human agency.⁴⁰

References

- Albritton, R.: 1995, 'Comments on Moore's Paradox and Self-Knowledge', *Philosophical Studies* 77, 229-239.
- Armstrong, D.: 1968, *A Materialist Theory of Mind*, Routledge & Kegan Paul, London.
- Bach, K. and R. Harnish: 1979, *Linguistic Communication and Speech Acts*, MIT Press, Cambridge (MA.).
- Bernecker, S.: 1996, 'Externalism and the Attitudinal Component of Self-Knowledge', *Nous* 30, 262-275.
- Bilgrami, A.: 1992, *Belief and Meaning*, Basil Blackwell, Oxford.
- Bilgrami, A.: 1998, 'Self-Knowledge and Resentment', in Cr. Wright, B. Smith and C. MacDonald (eds.), *Knowing Our Own Minds*, Clarendon Press, Oxford, 206-241.
- Boghossian, P.: 1989, 'Content and Self-Knowledge', *Philosophical Topics* 17, 5-26.
- Boghossian, P.: 1994, 'The Transparency of Mental Content', *Philosophical Perspectives* 8, 33-50.
- Brueckner, A.: 1990, 'Scepticism about Knowledge of Content', *Mind* 99, 447-451.
- Brueckner, A.: 1994, 'Knowledge of Content and Knowledge of the World' *Philosophical Review* 103, 327-343.
- Burge, T.: 1982, 'Other Bodies', in A. Woodfield (ed.), *Thought and Object*, Clarendon Press, Oxford, 97-119.
- Burge, T.: 1986, 'Intellectual Norms and the Foundation of the Mental', *Journal of Philosophy* 83, 697-720.
- Burge, T.: 1988, 'Individualism and Self-Knowledge', *Journal of Philosophy* 85, 649-663.
- Burge, T.: 1989, 'Wherein is Language Social?', in A. George (ed.), *Reflections on Chomsky*, Basil Blackwell, Oxford, 175-191.
- Burge, T.: 1993, 'Content Preservation', *Philosophical Review* 102, 457-488.
- Burge, T.: 1996, 'Our Entitlement to Self-Knowledge', *Proceedings of the Aristotelian Society* 117, 91-116.
- Burge, T.: 1998, 'Memory and Self-Knowledge', in Peter Ludlow and Norah Martin (eds.), *Self-Knowledge and Externalism*, CSLI Publication, Stanford, 351-370.
- Davidson, D.: 1984a, *Inquiries into Truth and Interpretation*, Clarendon Press, Oxford.
- Davidson, D.: 1984b, 'First Person Authority', *Dialectica* 38, 101-111.
- Davidson, D.: 1987, 'Knowing One's Own Mind', *Proceedings and Addresses of the APA* LX, 441-458.
- Davidson, D.: 1986, 'Rational Animals', in E. Lepore (ed.), *Truth and Interpretation*, Basil Blackwell,

Oxford, 473-480.

Davidson, D.: 1990, 'The Structure and Content of Truth', *Journal of Philosophy* 87, 279-328.

Davidson, D.: 1991, 'Subjektiv, Intersubjektiv, Objektiv', *Merkur* 45, 999-1014.

Davidson, D.: 1992, 'The Second Person', *The Wittgenstein Legacy, Midwest Studies in Philosophy XVII*, edited by P. French et. al., Notre Dame, University of Notre Dame Press.

Evans, G.: 1982, *Varieties of Reference*, Clarendon Press, Oxford.

Falvey, K. and J. Owens: 1994, 'Externalism, Self-Knowledge and Skepticism', *Philosophical Review* 103, 107-137.

Fodor, J.: 1992, *A Theory of Content and Other Essays*, MIT Press, Cambridge (MA).

Goldberg, S.:1999, 'The Relevance of Discriminatory Knowledge of Content', *Pacific Philosophical Quarterly* 80, 136-156.

Heil, J.: 1988, 'Privileged Access' *Mind* 47, 238-251.

Heil, J. 1992a: The Nature of True Minds, Cambridge, Cambridge University Press.

Heil, J.: 1992b, 'Believing Reasonably', *Nous* 26, 47-62.

Hymers, M. 1997: "Realism and Self-Knowledge: A Problem for Burge," in Philosophical Studies 86, pp. 303-325.

Jacobsen, R.: 1996, 'Wittgenstein on Self-Knowledge and Self-Expression', *Philosophical Quarterly* 46, 12-30.

H.H. Kögler and K. R. Stueber (eds.): 2000, *Empathy and Agency: The Problem of Understanding in the Human Sciences*, Westview Press, Boulder.

Ludlow, P./Martin, N. (eds.): 1998, *Self-Knowledge and Externalism*, CSLI Publication, Stanford.

Lyons, W.: 1986, *The Disappearance of Introspection*, MIT Press, Cambridge(MA).

McDowell, J.:1980, 'Meaning, Communication and Knowledge', in Z. Van Straaten (ed.), *Philosophical Subject*, Oxford University Press, Oxford , 117 - 139

Mele, A.: 1987, *Irrationality*, Oxford University Press, Oxford.

Mellor, D. H.: 1978, 'Conscious Beliefs', *Proceedings of the Aristotelian Society*, 87-101.

Millikan, R.: 1986, 'The Price of Correspondence Theory', *Nous* 20, 453-468.

Millikan, R.: 1993, *White Queen Psychology*, MIT Press, Cambridge (MA).

Nisbett, R. and T. Wilson: 1977, 'Telling More Than We Can Know: Verbal Reports on Mental Processes', *Psychological Review* 84, 231-259.

- Peacocke, Chr.: 1996, 'Entitlement, Self-Knowledge and Conceptual Redeployment', *Proceedings of the Aristotelian Society* 117, 117-158.
- Perry, J.: 1993, *The Problem of the Essential Indexical*, Oxford University Press, Oxford.
- Putnam, H.: 1988, Representation and Reality, MIT Press, Cambridge (MA.).
- Rosenthal, D.: 1995, 'Moore's Paradox and Consciousness', *Philosophical Perspectives* 9, 312-333.
- Shoemaker, S.: 1988, 'On Knowing One's Own Mind', *Philosophical Perspectives* 2, 183-209.
- Shoemaker, S.: 1991, 'Rationality and Self-Consciousness', in K. Lehrer and E. Sosa (eds.), *The Opened Curtain. A US-Soviet Philosophy Summit*, Westview Press, Boulder, 127-149.
- Shoemaker, S.: 1994, 'Self-Knowledge and Inner Sense', *Philosophy and Phenomenological Research* 54, 249-314.
- Shoemaker, S.: 1996, *The First Person Perspective and Other Essays*, Cambridge University Press, Cambridge.
- Stueber, K.: 1993, *Donald Davidsons Theorie sprachlichen Verstehens*, Anton Hain, Frankfurt a.M.
- Stueber, K.: 1997a, 'Holism and Radical Interpretation', in Analyomen 2, ed. G. Meggle, DeGruyter, Berlin/New York.
- Stueber, K.: 1997b, 'Psychologische Erklärungen im Spannungsfeld des Interpretationismus und Reduktionismus', *Philosophische Rundschau* 44, 304-328.
- Stueber, K.: 2000, 'Understanding Other Minds and the Problem of Rationality', in Hans Herbert Kögler and Karsten R. Stueber (eds.), *Empathy and Agency: The Problem of Understanding in the Human Sciences*, Westview Press, Boulder.
- Tugendhat, E.: 1976, *Vorlesungen zur Einführung in die sprachanalytische Philosophie*, Suhrkamp, Frankfurt a.M..
- Wright, Cr.: 1991, 'Wittgenstein's Later Philosophy of Mind: Sensation, Privacy and Intention', in K. Puhl (ed), *Meaning Scepticism*, de Gruyter, Berlin/ New York, 126-147.
- Wright, Cr.: 1989, 'Wittgenstein's Rule-Following Consideration and the Central Project of Theoretical Linguistics', in A. George (ed.), *Reflections on Chomsky*, Basil Blackwell, Oxford, 233-264.
- Wright, Cr., B. Smith, and C. MacDonald : 1998, *Knowing Our Own Minds*, Clarendon Press, Clarendon Press.

Endnotes

-
1. For a short history of the introspective paradigm in psychology consult W. Lyons 1986. For an example of current psychological research demonstrating our "ignorance" about mental processes see R. Nisbett and T. Wilson 1977.
 2. This is the route that R. Millikan is inclined to take. According to her externalist and naturalistic account of semantic content there seems to be no room for the assumption that speakers know the truth-conditions of their own thoughts in a privileged manner. See for example Millikan 1993, p.283 and 1986, p.465.
 3. This is not to deny that in certain circumstances my knowledge of my mental states can be based on such evidence.
 4. By calling self-knowledge "non-empirical" I am only referring to its epistemic status, not its "object." I am not denying that self-knowledge is about mental facts in the world. I am only emphasizing that it does not require further empirical evidence. Although this characterization of self-knowledge makes it certainly tempting to think of it on the model of the perception of an external object, I will shortly indicate at the beginning of the first section why I regard the perceptual model as unpromising.
 5. For the purpose of this paper, I will focus only on the causal form of externalism and disregard Burge's social externalism. First of all, I do not think that social externalism introduces a different type of problem as far as self-knowledge is concerned. Secondly, in his later publication Burge himself seems to conceive of causal externalism as the more fundamental form, since he derives the social dimension of language use from it. See Burge 1989 and 1986. In this paper I will also not worry about the question of whether or not Davidson's form of externalism is compatible with Burge and Putnam's exposition of it. In an interesting paper, Hymers (1997) argues that Burge's externalism in contrast to Davidson's is committed to the thesis of metaphysical realism according to which our theories of the world could be completely wrong. He furthermore shows that such externalism is in principle incompatible with the assumption of self-knowledge. However, I do not find it necessary to take a stance on whether Burge himself is indeed committed to the thesis of metaphysical realism. I would argue that twin earth thought experiments provide a plausible intuition pump for externalism even from the perspective of a more modest form of realism, according to which the world exist independently of the mind but the mind cannot have a completely mistaken picture of the world. In order to get these thought experiments off the ground, it is only required that we could be wrong about the microstructure of certain natural kinds.
 6. Notice, however, that I emphasize the process of deliberation. While first order beliefs seem to be sufficient to rationalize actions, the capacity for reflective deliberation does not seem intelligible without the assumption of self-knowledge. I, however, do not regard this insight sufficient to justify the assumption of self-knowledge, as will be shown later.
 7. See J. Perry "The Problem of the Essential Indexical" in his 1993.
 8. For a detailed and lucid discussion of the object perception model see S. Shoemaker 1994, esp. Lecture I.
 9. Since I am not primarily interested in interpreting Wittgenstein in this paper, I will not address the question of whether Wittgenstein is committed to a mere expressivist account of self-avowals as is customarily assumed. First person authority would then be established by denying that self-ascriptions are self-reports and thus the expression of real self-knowledge. Such position would seem to violate the asymmetry constraint. For a balanced explication of Wittgenstein's expressionism see R. Jacobsen 1996. Crispin Wright's neo-Wittgensteinian analysis of first person authority seems to avoid certain problems of Wittgenstein. Wright follows Wittgenstein in denying that first person authority depends on a "substantial epistemology" or specific cognitive mechanism, but he does not follow him in the expressivist analysis of self-ascriptions. He conceives of self-knowledge as an essential part of our notion of intentionality and our interpretative practice. See especially Cr. Wright 1991, p. 142 and 1989, pp. 251/52. Nevertheless even Wright's constitutive account of self-knowledge begs the question against the skeptical worries raised in the context of externalism.

10. A. Brueckner 1990, p.448. For further specifications of this skeptical challenge see Brueckner 1994 and P. Boghossian, 1989. In this paper I am focusing on what one can call with Davies the achievement problem of self-knowledge, given externalism. (Davies "Externalism, Architecturalism and Epistemic Warrant" in Wright et. al 1998, p. 340.) It has to be distinguished from the consequence problem of self-knowledge, given externalism. According to the consequence problem the acceptance of self-knowledge and externalism leads to the absurd consequence that one can know facts about the world in a non-empirical way. See for example McKinsey and Brown (Reprinted in Ludlow/Martin 1998). I tend to agree with McLaughlin/Tye and Raffman (in Wright et. al. 1998) that this argument rests on a confusion about the theses to which the externalist has committed himself. But this has to be spelled out in another paper.

11. T. Burge 1988, and especially J. Heil 1988 and 1992, chap.5.

12. See Brueckner 1990, p.449.

13. Burge 1988, pp. 658/59. See also pp. 654/56. Burge tries to provide a more general explication of our entitlement to self-knowledge in Burge 1996. For a short critique of that account see the end of the next section.

14. See Brueckner 1990, p. 451, fnote. 11.

15. Falvey and Owens seem to hint at that response without developing the argument and recognizing its central significance in the strategy against the Brueckner's skeptic. See their 1994, p.120. I assume here, that it is coherent to entertain simultaneously a hypothesis about water and twin-water thoughts in the context of externalism. In my opinion all that is required to raise worries regarding twin-water thoughts is the abstract possibility that the external constitution of our world might be radically different, and such skeptical worries seem to be perfectly intelligible.

16. For an elaboration of this line of thought see P. Boghossian, 1994, p.38f. In his 1989, p.23 Boghossian uses a similar argument about slow-switching in order to argue that externalism is incompatible with self-knowledge of one's occurrent thought, not only with the comparative transparency of thought. For a discussion of this argument and its problematic assumptions about the nature of memory, see especially the articles by Ludlow, Brueckner and Bernecker in Ludlow/Martin 1998. Burge 1998 (p.364f) argues against Boghossian that lack of epistemic transparency in the situation of the switching cases does not automatically imply a lack of rationality and invalid forms of reasoning. According to Burge, such a conclusion can be avoided if one recognizes the "centrality of preservative memory" (366) in reasoning. Preservative memory contains the original conceptual framework and allows one to avoid equivocation in reasoning. Boghossian's examples are for him only cases of memory misidentification and not failures of reasoning. However, I do not find this answer to Boghossian completely satisfactory, since it still seems to allow for such cases of memory misidentification to be rather pervasive. Rationality is then preserved only at the price of the agent not knowing what he is actually reasoning about. Such a notion of rationality seems to deviate from our ordinary notion of agent rationality, according to which reasons are not only located in an agent but are also reasons for an agent. However as I will argue later on, Boghossian's argument does not succeed in showing that lack of epistemic transparency is a widespread phenomenon. For a defense of Boghossian's concerns see also Goldberg 1999.

17. John Heil emphasized this point in a written correspondence. See also Bernecker 1996.

18. Davidson 1987, p.456 and 1984b, p.111. In his 1998, Akeel Bilgrami develops a very interesting account of self-knowledge according to which self-knowledge is a necessary condition for holding agents responsible. For Bilgrami, self-knowledge itself is therefore a normative notion since it cannot be accounted for independent of our normative reactive attitudes. I am very sympathetic to Bilgrami's constitutive account and share his reservations against a purely reliabilist account of self-knowledge. I see self-knowledge, however, as a constitutive assumption of the more basic practice of attributing meaning and intentional states to other persons.

19. Compare also the illuminating article by Barry Smith (1998). Smith, like me, claims that Davidson fails to answer the question of how we know our own minds given that we do not interpret ourselves (p.409). In contrast to me, Smith understands this question in light of Dummett's critique of Davidson's theory of meaning. He demands that in order to answer this question one has to explain how we know from the first person perspective what we mean and what our

knowledge of truth-conditions consist in. For my understanding of a Davidsonian response to Dummett see my 1993. I do not think that an explanatory answer to this question can be given, since I do not believe that there is an interesting distinction between grasping the meaning of a word and using it correctly.

20. Shoemaker 1996, p.27.

21. *ibid.*, p. 34.

22. *ibid.*, p. 236.

23. See G. Evans 1982, p.225/26. For further critique of Shoemaker see Albritton 1995, pp.231-35 and Rosenthal 1995, p. 327.

24. For a very plausible taxonomy of different speech acts and for a plausible theoretical explication of communication see K.Bach and R. M. Harnish (1979).

25. This is an objection Thomas Grundmann raised. Rosenthal might argue similarly. He derives the assumption of self-knowledge from the claim that "any intentional state that I express with a speech act will be a conscious intentional state." Rosenthal 1995, p. 317. For Rosenthal, a belief that *p* is conscious only if I have a second order belief about it. Insofar as a sincere assertion requires a conscious belief it also requires a second order belief. However, he counts only sincere assertion as a genuine speech act, since insincere assertions do not express conscious beliefs. His claim that one cannot succeed in making a genuine assertion in this case seems to be forced on us by merely theoretical considerations, which are not based on plausible linguistic considerations.

26. Davidson (1990, 310/11 and 1992, 258) requires that one interprets the speaker as he "intended to be interpreted," without directly stating what such an intention entails for very different kinds of speech acts. I take my account to be a more precise characterization of such a very broad intention insofar as the speech act of an assertion is concerned. Without such a detailed analysis it is difficult to show why the assumption of self-knowledge is a constitutive assumption of linguistic agency. I also do not think of intentions to express certain mental attitude as being necessarily the outcome of a process of deliberation. In contrast to my account, J. McDowell (1980) and E. Tugendhat (1976, lecture 14 & 15) claim that an assertion has to be understood in light of an intention to say something true. However, McDowell's and Tugendhat's account of an assertion remains unsatisfactory, since they do not explicate what constitutes the speech act of saying that something is true or how we can explicate the intention to say something true. Equally important, I am not sure how one could account for lying as asserting something within this conception of assertion.

27. In his 1978 (p.96ff), D. H. Mellor uses a Gricean account of communication in order to argue that the existence of second order beliefs is essential for linguistic action. I am however not sure, how he can account for the possibility of lying within this context, as he claims is possible. More importantly and in contrast to my position, Mellor accounts for the truth-conduciveness of these second order beliefs causally and not conceptually. For him, self-knowledge is based on a reliable mechanism of the brain. For some brief remarks on the compatibility of such a reliabilist account of self-knowledge and my constitutive account see the conclusion of this paper.

28. For the purposes of this paper, I do not have to decide what is the philosophically best account of voluntary agency and free will. Rather, I appeal to common folk-psychological assumptions about linguistic agency and argue that they imply that we have to attribute self-knowledge to linguistic agents.

29. For a critical discussion of the exact nature of this rationality assumption see my 2000. It would be certainly too strong to require that agents are ideally rational as articulated by our formal theories of reasoning.

30. For a detailed analysis of Davidson's notion of radical interpretation see my 1993. For a defense of radical interpretation against Fodor's and Chomsky's objections see my (1997a).

31. This paragraph is due to concerns, which Gary Ebbs raised in his comments to my paper at a meeting of the Society for Philosophy and Psychology. Ebbs' concerns articulate a worry that I have encountered quite often. One seems to

wonder of how one can accept certain features of Grice's theory within the Davidsonian framework. More specifically one asks how one can assume that a speaker has intentions that are conceived of as the standards for interpretation and at the same time maintain that meaning and mental content are constituted by the evidence for an interpretive theory of truth as revealed by the analysis of radical interpretation. This worry rests on a misunderstanding of Davidson's theory. Davidson never objected to Grice's positions as an analysis of the complicated relation between meaning and intentions. He objected only against understanding Grice as providing a reductive account of linguistic meaning, i.e. as taking these intentions as primary bearer of content in comparison to linguistic meaning, since we are able to attribute meaning and mental content only simultaneously. See for example Davidson "Belief and the Basis of Meaning," in his 1984a, p.143. However, a constitutive account of self-knowledge, as I outline it, does not square very well with Davidson's doctrine of triangulation, i.e. the idea that without actual communication between two people it does not make sense to speak of the content of a person's mental states. Since in interpreting another person we have to assume that she knows her own thoughts we seem to be committed to presuppose self-knowledge even before the act of expressing a specific belief. This implies that actual communication is not constitutive for one's thoughts having a certain content. I would therefore suggest that Davidson should give up on triangulation. For Davidson's thesis of triangulation see his 1991 and 1986.

32. See the articles by Shoemaker listed in the bibliography and Burge 1996. For a critique of Burge see also Peacocke 1996.

33. Burge 1993, 458.

34. See Shoemaker 1988, p.33.

35. For a response of Burge against Boghossian see however footnote 16. A thorough discussion of Burge's notion of preservative memory would require another article. Here it should be noted that a Davidsonian response to Boghossian does not require appeal to such a concept.

36. The thought that one could re-experience the past objectively, was an idea closely associated with the suggestion that empathy is the central method of the human sciences. For a comprehensive survey of the debate about empathy (including the contemporary notion of simulation) see the introduction ("Empathy, Simulation and Interpretation in the Philosophy of Social Science) of Kögler/Stueber 2000. The various chapters of this anthology evaluate the contemporary revival of empathy as proposed by simulation theorists from various perspectives in the philosophy of social science, philosophy of mind and interpretation theory.

37. For the different strategies of self-deception see A. Mele 1987, chap.10. I do not want to deny that judging the rationality of belief acquisition can be a difficult task because one not only has to take into account epistemic norms but one has to see the behavior in light of the non-epistemic interest of a particular agent. In this case however I do not believe that one can interpret the actions of the mother as being rational even on a broader conception of rationality. Assuming that the mother loves her son and wants to preserve his well being, it is not in her best practical interest to not address the self-destructive behavior of her son. See J. Heil (1992b) for the non-epistemic aspect of rationality.

38. D. Armstrong 1968, especially p.326. I.e. he does understand scanning as the perception of an object. Lack of self-knowledge could then be explained as the malfunctioning of the scanning mechanism, i.e. a mistake in the scanning process occurs if a second order state is caused by something other than the first order state which **normally** causes it to be activated. (I here assume, of course, that the normalcy condition can be defined in a non-circular way. The normalcy condition is required since the scanning model would have the same problems to account for error as any information based semantics. For a short overview see J. Fodor 1992, pp. 51-87.)

39. See especially H. Putnam 1988. I critically discuss such naturalistic accounts of content – especially Dretske and Millikan's teleosemantic versions - in my 1997b.

40. I would like to thank Sven Bernecker, Gary Ebbs, Christopher Dustin, Sandy Goldberg, Thomas Grundmann, John Heil, Marya Schectman, Manisha Sinha, and an anonymous referee of the journal for their valuable suggestions on earlier versions of this article.